

WHAT IS CLAIMED IS:

1. A method for obtaining DNA fingerprint profile data, comprising the steps of:

measuring peak intensity and size of each DNA fragment in a sample of genomic DNA;

classifying the peak intensities of said fragments according to a predetermined discrete intensity level scale;

aligning the sizes of said fragments into
corresponding ones of discrete size bins;

entering the classified peak intensities of said fragments into a data record in a sequence determined by said aligned sizes; and

storing the record.

2. The method of claim 1, wherein said sample is obtained by ALFP.

3. The method of claim 1, wherein said sample is obtained by RLFP.

4. The method of claim 1, wherein said sample is obtained by SSR PCR.

5. The method of claim 1, wherein said sample is obtained by VNTR PCR.

6. The method of claim 1, further comprising the step of normalizing the measured peak intensities of said fragments before classifying said peak intensities according to said discrete intensity level scale.

7. The method of claim 6, wherein said discrete intensity level scale comprises at least five discrete peak levels.

8. The method of claim 1, wherein the step of aligning comprises the steps of:

grouping fragment sequences of said fragments into clusters, wherein adjacent peaks of each cluster are less

Sub A1

	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100	2101	2102	2103	2104	2105	2106	2107	2108	2109	2110	2111	2112	2113	2114	2115	2116	2117	2118	2119	2120	2121	2122	2123	2124	2125	2126	2127	2128	2129	2130	2131	2132	2133	2134	2135	2136	2137	2138	2139	2140	2141	2142	2143	2144	2145	2146	2147	2148	2149	2150	2151	2152	2153	2154	2155	2156	2157	2158	2159	2160	2161	2162	2163	2164	2165	2166	2167	2168	2169	2170	2171	2172	2173	2174	2175	2176	2177	2178	2179	2180	2181	2182	2183	2184	2185	2186	2187	2188	2189	2190	2191	2192	2193	2194	2195	2196	2197	2198	2199	2200	2201	2202	2203	2204	2205	2206	2207	2208	2209	2210	2211	2212	2213	2214	2215	2216	2217	2218	2219	2220	2221	2222	2223	2224	2225	2226	2227	2228	2229	2230	2231	2232	2233	2234	2235	2236	2237	2238	2239	2240	2241	2242	2243	2244	2245	2246	2247	2248	2249	2250	2251	2252	2253	2254	2255	2256	2257	2258	2259	2260	2261	2262	2263	2264	2265	2266	2267	2268	2269	2270	2271	2272	2273	2274	2275	2276	2277	2278	2279	2280	2281	2282	2283	2284	2285	2286	2287	2288	2289	2290	2291	2292	2293	2294	2295	2296	2297	2298	2299	2300	2301	2302	2303	2304	2305	2306	2307	2308	2309	2310	2311	2312	2313	2314	2315	2316	2317	2318	2319	2320	2321	2322	2323	2324	2325	2326	2327	2328	2329	2330	2331	2332	2333	2334	2335	2336	2337	2338	2339	2340	2341	2342	2343	2344	2345	2346	2347	2348	2349	2350	2351	2352	2353	2354	2355	2356	2357	2358	2359	2360	2361	2362	2363	2364	2365	2366	2367	2368	2369	2370	2371	2372	2373	2374	2375	2376	2377	2378	2379	2380	2381	2382	2383	2384	2385	2386	2387	2388	2389	2390	2391	2392	2393	2394	2395	2396	2397	2398	2399	2400	2401	2402	2403	2404	2405	2406	2407	2408	2409	2410	2411	2412	2413	2414	2415	2416	2417	2418	2419	2420	2421	2422	2423	2424	2425	2426	2427	2428	2429	2430	2431	2432	2
--	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	---

5 than or equal to a predetermined number of discrete positions apart;

assigning a potential energy value to each cluster, said potential energy value being proportional to the spacing between adjacent peaks of the cluster and to the
10 amount of displacement required to bin the peaks into discrete size bins;

varying the displacement of said peaks so that said potential energy value is minimized; and

aligning said peaks into discrete size bins according to the displacement values that result in a minimized potential energy value.
15

9. The method of claim 1, wherein said step of entering comprises the step of creating a data record having: an information field including

an identification of the source of the genomic DNA,
5 the method of producing DNA fragments from said source,

the spacing between successive fragments,
the starting size of said fragments, and
the ending size of said fragments; and

10 a sequence field containing a sequence of said classified peak intensities.

10. The method of claim 9, further comprising the step of scoring a comparison between two data records, including the steps of assigning reward points to matches of peak intensities and penalty points to mismatches of peak intensities, based on the relative magnitude of said peak intensities, comparing said two data records for identical matches at corresponding size positions, comparing said two data records for identical matches at adjacent size positions, comparing said two data records for non-

identical matches at corresponding size positions, and comparing said two data records for non-identical matches at adjacent size positions, and totaling said reward points and penalty points according to found matches and remaining mismatches to obtain a score.

11. The method of claim 10, further comprising the step of obtaining a percentage metric for two compared data records, by obtaining a ratio of said score to a score obtained by matching one of said two data records to itself.

12. A computer program product, comprising:

a computer-readable medium having computer-executable code recorded thereon for obtaining DNA fingerprint profile data, said computer-executable code comprising:

5 means for measuring peak intensity and size of each
DNA fragment in a sample of genomic DNA;

means for classifying the peak intensities of said fragments according to a predetermined discrete intensity level scale;

10 means for aligning the sizes of said fragments into
corresponding ones of discrete size bins;

means for entering the classified peak intensities of said fragments into a data record in a sequence determined by said aligned sizes; and

15 means storing the record in a computer-readable
storage medium.

13. The computer program product of claim 12, further comprising:

means for normalizing the measured peak intensities of said fragments before classifying said peak intensities according to said discrete intensity level scale.

14. The computer program product of claim 13, wherein said discrete intensity level scale comprises at least five discrete peak levels.

15. The computer program product of claim 12, wherein said means for aligning comprises:

means for grouping fragment sequences of said fragments into clusters, wherein adjacent peaks of each cluster are less than or equal to a predetermined number of discrete positions apart;

means for assigning a potential energy value to each cluster, said potential energy value being proportional to the spacing between adjacent peaks of the cluster and to the amount of displacement required to bin the peaks into discrete size bins;

means for varying the displacement of said peaks so that said potential energy value is minimized; and

means for aligning said peaks into discrete size bins according to the displacement values that result in a minimized potential energy value.

16. The computer program product of claim 12, wherein said means for entering comprises means for creating a data record having:

an information field including

an identification of the source of the genomic DNA, the method of producing DNA fragments from said source,

the spacing between successive fragments,

the starting size of said fragments, and

the ending size of said fragments; and

a sequence field containing a sequence of said classified peak intensities.

17. A DNA fingerprint data record stored in a computer-readable storage medium, said DNA fingerprint data including size and peak intensity of DNA fragments produced from a DNA sample obtained from a source of genomic DNA, said data record comprising:

an information field including

an identification of said source of genomic DNA,

the method of producing DNA fragments from said source,

the spacing between successive fragments,

the starting size of said fragments, and

the ending size of said fragments; and

a sequence field containing a sequence of classified peak intensities.